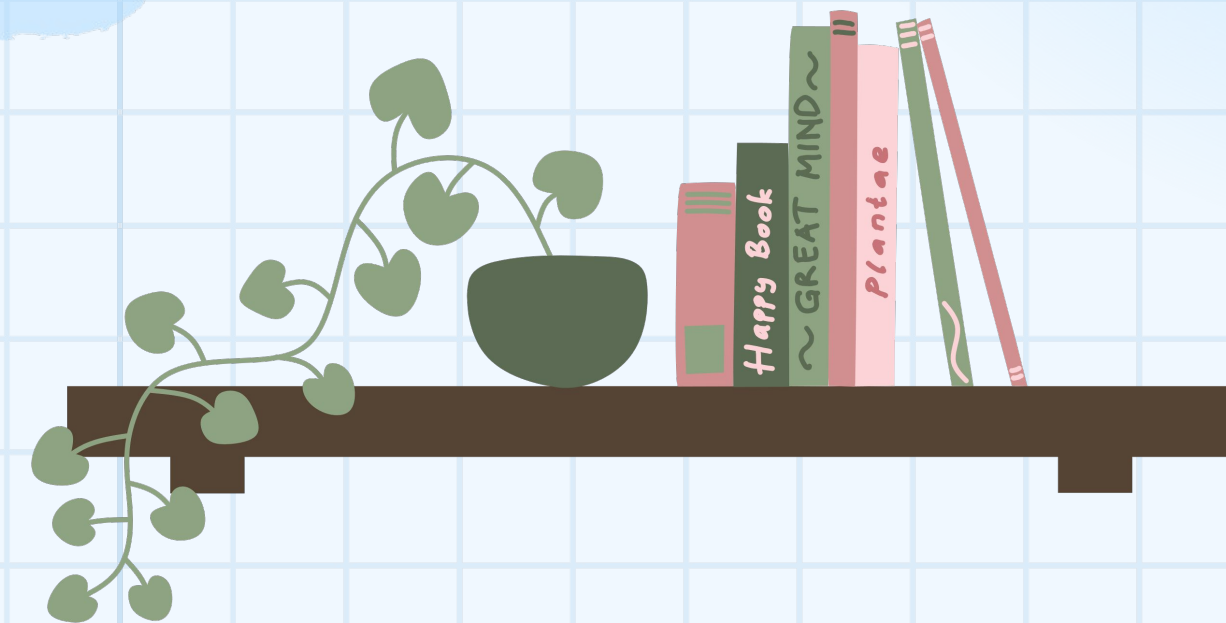
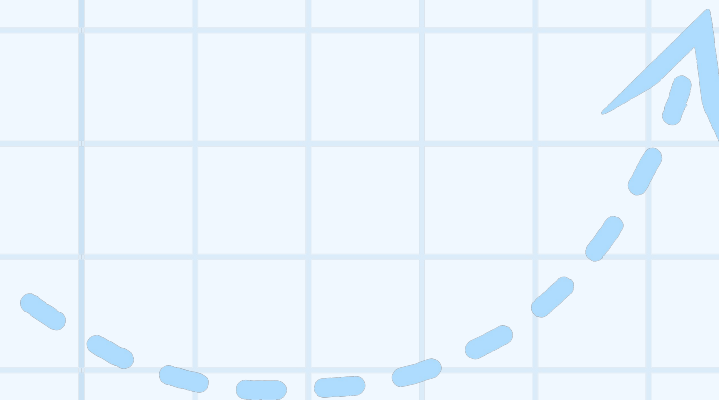
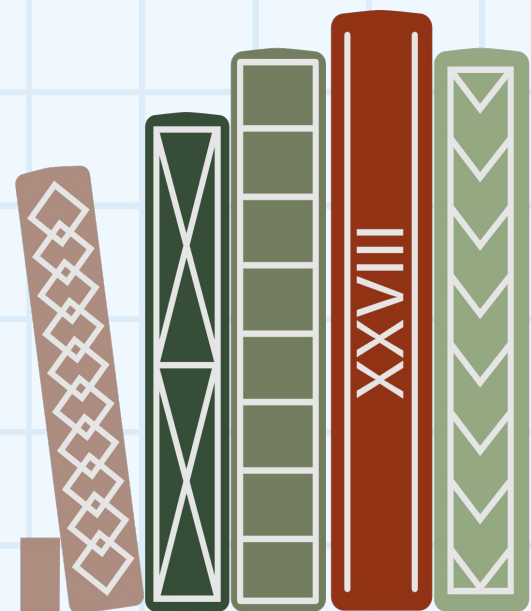




# BS./BSC.IN

Applied ai and Data Science

## Basics of Data Analytics



# Let's dive into and learn:



1

## Missing Data

# Missing Data



Powered by



**FutureSense**

- After checking for duplicates and formatting errors, check for missing data or missing values
- Missing values are very common
- Missing values can occur for various reasons
- Requires handling carefully

# Missing Data



Powered by



**FutureSense**

- Not have information for some records
  - Some users did not answer the question
  - Question not applicable to a particular user
- Information did not get recorded
  - Machine malfunction
- Many other reasons for missing values





# Handling Missing Data

- Two options for dealing with missing data
  - Filling in the missing values
    - Can be a preferred method in many cases
    - Doesn't mean you invent a value!
  - Delete the records altogether
    - Necessary to do sometimes
    - Be careful

# Handling Missing Data



- Filling in the missing values
  - Fill in with an indicator of missingness e.g. NA or N/A
  - Do the same for every case of missing value
  - Carry on analysis



# Handling Missing Data

Transaction ID	Customer ID	Product ID	Purchase Date	Customer Name	Location
53666	34987	453	01-02-2025	Priya	Patna
54333	35671	984	12-01-2025	Rahul	
56473	39823	382	29-01-2025	Divesh	Bhopal
56234	36005	226	05-01-2025	Neha	Pune

Can replace missing values with an indicator for missingness and carry on analysis



Transaction ID	Customer ID	Product ID	Purchase Date	Customer Name	Location
53666	34987	453	01-02-2025	Priya	Patna
54333	35671	984	12-01-2025	Rahul	NA
56473	39823	382	29-01-2025	Divesh	Bhopal
56234	36005	226	05-01-2025	Neha	Pune



# Handling Missing Data

- Delete entire observations
  - When crucial information is missing e.g. the information you want to analyze, some unique identifier
  - Could potentially bias the data and lead you to draw the wrong conclusions, E.g. suppose data from a particular store have missing values



# Handling Missing Data



Transaction ID	Customer ID	Product ID	Purchase Date	Customer Name	Location
53666	34987	453	01-02-2025	Priya	Patna
54333	35671	984	12-01-2025	Rahul	Bikaner
56473	39823		29-01-2025	Divesh	Bhopal
56234	36005	226	05-01-2025	Neha	Pune



When crucial information is missing  
need to delete entire observations

Transaction ID	Customer ID	Product ID	Purchase Date	Customer Name	Location
53666	34987	453	01-02-2025	Priya	Patna
54333	35671	984	12-01-2025	Rahul	Bikaner
56234	36005	226	05-01-2025	Neha	Pune



# Identifying Missing Data

- Sometimes missing information may appear in the data other than blanks
  - N.A. or some variation of it
  - .
  - -
  - -99
  - -100 and many more

# Recap



## Handling Missing Data





# Thank you

